Analytics for tracking student engagement

Christine Gardner

Open University, UK
c.f.gardner@open.ac.uk

Helen Jefferis

Open University, UK
h.jefferis@open.ac.uk

Allan Jones

Open University, UK
allan.jones@open.ac.uk

Abstract

Although there has been much research in the area of data analytics in recent years (e.g. Shum and Ferguson 2012), there are questions regarding which analytic methodologies can be most effective in informing higher education teaching and learning practices (Gibson and de Freitas, 2016).
This project focuses on one module within the School of Computing and Communications in the STEM faculty to gain a clearer understanding on why students might, or might not, engage with computer aided learning and teaching (TELT) resources. We explore the use of specific TELT resources on the module 'Communications Technology', a print-based module with a range of online resources designed to supplement the text.

The research questions cover two key areas; the effectiveness of the analytics tools and students' perception of the TELT resources.

Via data analytics we can review:

- When the students engage with the TELT resources and whether this is at predicted times during the module.
- Whether students revisit the TELT resources.

Via individual student feedback we can explore:

- What motivates students to engage with TELT resources.
- Whether students understand topic more deeply as a result of using TELT resources.
- If students are deterred if the resources are too complicated/time consuming.

The findings should be of interest to module teams across many universities. This project will build on previous work undertaken in this area, e.g. Herodotou et al (2017) and Tempelaar et al (2017), and contribute to the wider body of knowledge in the area of data analytics.

Keywords: data analytics, informatics

1. Introduction

The Open University (OU) has evolved significantly since its creation fifty years ago and has developed its own style of distance learning, 'supported open learning', offering students opportunities to study flexibly, whether at home, work, library or other study centre. Before the advent of the Internet, students relied solely on printed study materials. Key to its continuing success is the utilisation of new technologies.

This study looks at the use of learning analytics to uncover student engagement with computer aided learning and teaching (TELT) resources in the UK Open University module TM355 Communcations Technology. This module is an elective component in the University’s honours degree in Computing and IT. The module covers such topics as radio propagation, digital signal modulation, source coding, error control, optical fibres, DSL broadband and mobile communications. Parts of the module are supported by sophisticated TELT resources, particularly in relation to coding and error control.

The module is studied towards the end of the students’ degree level studies and introduces several complex topics. To aid study of such material, additional experiential learning (Kolb, 1984) is available via online interactive activities, designed to supplement the written materials. These are referred to within the printed materials and are added to the students’ study planner, grouped together to make them relatively easy to find. An example is shown in Figure 1 below.

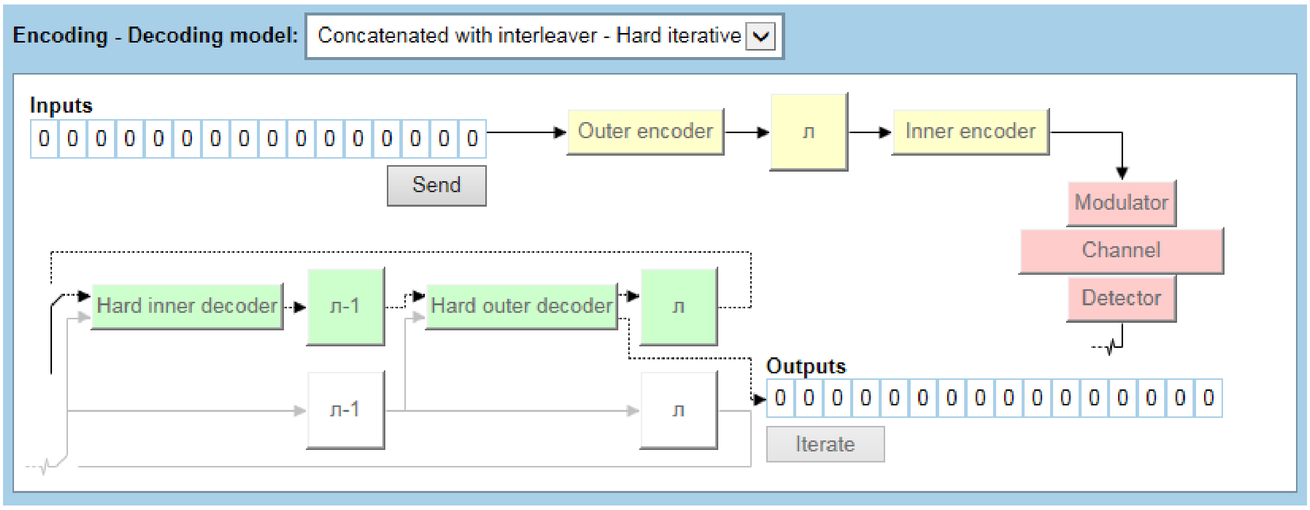

Figure 1 – Example of TM355 TELT resource (online activity)

The research was motivated by a particular examination question used in the 2017 examination. The question, on the topic of error control, was in a part of the examination paper where students had a choice of questions to answer. The question related to techniques of error control that had been taught in print, and demonstrated interactively with a TELT resource which students were strongly advised to use, but could not be compelled to use. In this study learning-analytics data was used retrospectively to investigate the use of the relevant TELT resource by students who chose to answer the question, and as an aid to framing interview questions relating to the use of TELT resources. This appears to us to be a novel use of learning analytics data.

2. Learning analytics

Learning analytics, in George Siemens's widely used definition, are the 'measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs' (quoted in Bodily and Verbert, 2017, p. 405). The gathering and use of statistical data about learners is not new. For example, pass/fail rates and grade distributions have long been a tool of educators and educational researchers, but 'learning analytics' connotes something more than traditional performance statistics.

The development of learning analytics has largely been an outgrowth of the development of virtual learning environments (VLEs), where students' online study is conducted in a computer-based learning environment. Such an environment allows the students' progress, performance of tasks, use of resources, etc. to be recorded. Although such data could be used for monitoring an individual student, generally learning-analytics data is aggregated from many students (sometimes hundreds or thousands) in order to identify significant trends or patterns of study behaviour.

Although there has been much research in the area of data analytics in recent years (e.g. Shum and Ferguson 2012), there are questions regarding which analytic methodologies can be most effective in informing higher education teaching and learning practices (Gibson and de Freitas, 2016).
Indeed, learning analytics have also been seen as an outgrowth of the development of 'big data' movement. (Kop et al. 2017; Littlejohn 2017.) Two 'big data' applications of learning analytics have received particular attention. One is the use of learning analytics to predict students' behaviour or success. (for example, Slater and Baker 2019). The other is the use of learning analytics in conjunction with learning design to help with module revision and improvement. (Sclater et al. 2016). Slater and Baker (2019) point out a potential problem with the predictive approach, which is its tendency to assume that learning is continuous and incremental, allowing extrapolations to be made. This ignores the possibility of learning being discontinuous, in which the gaining of a sudden insight produces an unpredicted shift in performance. An additional problem with the 'predictive' approach is that although it can reveal correlations, causal connections between student activity and educational progress remain unclear. The problematic nature of predictive learning analytics possibly underlies a trend identified by Viberg et al. (2018, p. 108), who report that research in learning analytics in higher education is shifting from prediction towards 'a deeper understanding of students' learning experiences.' The authors of the present paper see their work as part of this trend. Nevertheless, in so far as the present study attempts to uncover a possible correlation between use of TELT resources and examination performance, it has elements of the 'predictive' approach and the 'student experience' approach.

Two issues in particular are common to both the 'predictive' approach and the 'student experience' approach. The first of these concerns the ethics of monitoring students. Siemens (2019) refers to concerns around student privacy in connection with learning analytics and Bodily & Verbert (2017) observe that some types of learning analytics potentially reduce student autonomy as teachers and administrators increasingly become framed as managers of learners.

The second issue concerns what analytics actually represent. Analytics data comprises counts of mouse-clicks made by students to get to particular virual learning environment (VLE) page or web pages, and, possibly, time spent on a particular web VLE page or web page. Records of clicks and time spent therefore serve as proxies for learning activities and resource use, and possibly as rather poor proxies. For instance, Macfadyen and Dawson (2010) find that time spent on educational resources, as indicated in learning analytics data, is poorly correlated with academic performance. Thus students' motivation for clicking on particular resources, and the degree of attention students pay to them, cannot reliably be deduced from analytics data. To investigate issues of motivation and attention supplementary techniques such as surveys and interviews must be used. In the present case, interview, learning analytics data was supplemented with interviews. This appears to the authors to be a novel. or at any rate uncommon, use of learning analytics.

3. Method

The research questions cover two key areas; the effectiveness of the analytics tool and students' perception of the TELT resources. The methodology employed a mix of quantitative and qualitative research methods, in particular the collection of data analytics and use of semi formal interviews. Via data analytics it was possibile to review when the students engaged with the TELT resources and whether at predicted times during the module. It was also possible to collect data to establish whether students revisit the TELT resources. Via individual student telephone interviews a more in-depth view could be established regarding what motivates students to engage with TELT resources, whether students understand topic more deeply as a result of using TELT resources, or if students are deterred if resources are too complicated or too time consuming.

3.1 Data collection - analytics for action (A4A)

The main data analytics tool selected for the research was Analytics for Action, A4A (Hidalgo, 2018). A4A can provide detail of how students are engaging with specific online materials. Data is presented at a high level, with the aim to provide a module-level analysis of how students are engaging with online materials. The framework for application of A4A has six phases, as shown in Figure 2, that can help module teams continually review and improve student experience by identifying specific actions to be taken.

Figure 2 – A4A six phases

A4A is a visual platform, providing a summary of student performance using real-time data. For example, Figure 3 depicts student interaction with a specific online resource that relates directly to assessed material. The vertical axis represents the number of students engaging with a TELT resource, the horizontal axis represents the study weeks of the module and thelight blue vertical bar represents an assignment due date. It can be seen that peak use of the related online resource ties in with the submission of the second assignment, due in week 20.

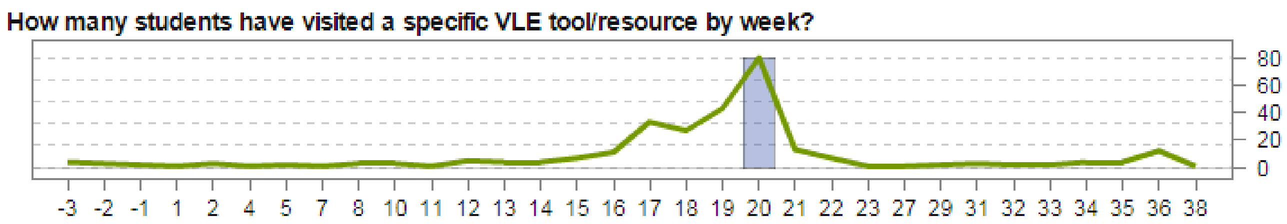

Figure 3 – Hamming codes - Online resource predominantly used in relation to assessment

In Figure 4, the use of the TELT resource has a different kind of pattern, as it is used predominantly during Block 2 of the module, between the first and second assignments (represented by the first and second vertical bars). There is also a small peak at the end of the module, suggesting that some students return to the online resource at revision time. However, the number of students engaging is relatively low, considering the cohort size of over 300 students.

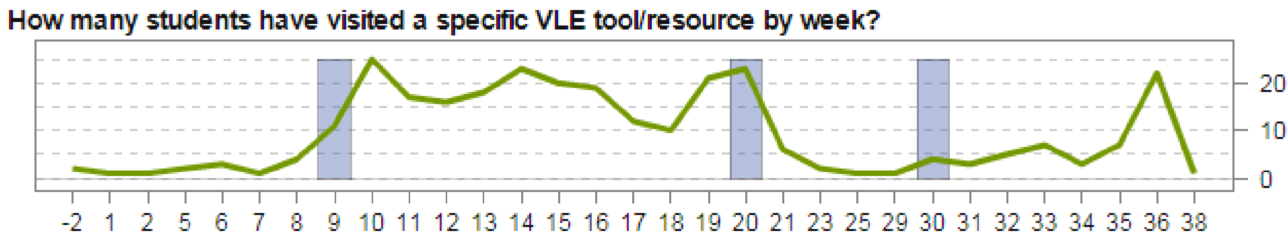

Figure 4 – Error control codes - Online resource used at specific times during a module

The A4A data can help a module team make evidence-based decisions, with the ultimate goal of improving student experience on that module (Evans et al, 2017). A limit to the usefulness of the A4A dashboard is that in its usual format it cannot identify online activity at an individual student level. In consultation with the analytics design team it was established that the underlying data could be presented at an individual student level if required, drilling down to explore the data in more depth.

3.2 Phases of research and ethical considerations

For this study, a sub-set of 48 students was selected, namely those who answered a specific TM355 2017 exam question on error correction. The question was not answered well resulting poor average performance. Student activity on the related TELT resource was collected and could then be mapped alongside their exam performance.

The research was conducted in three main phases. During the first phase (2017 to July 2018) a pilot study was conducted which commenced with the collection of key analytics data via A4A on TELT resource use during the 2016/17 module presentation. From this the sub-set of students was identified for further research. In consultation with analytics team this data was interrogated more deeply. The second phase was designed to supplement the analytics data via semi-formal interview questions, to help address limitations such as those noted by Macfadyen and Dawson (2010). Interviews took place July 2018 and was followed by an initial review of findings. The third phase involves action and dissemination of findings and is currently ongoing. The research was approved by the Open University's Research Ethics Committee and has been logged as GPDR compliant. The third phase involves action and dissemination of findings and is currently ongoing.

4. Results and discussion on recommendations

The following is a summary of findings from both the data analytics and student interviews.

4.1 Results from A4A regarding error coding activity relating to exam

In reviewing the data relating to the TM355 students who answered the specific exam question relating to error coding, the small sample of 48 students was selected from the cohort of 329 students who sat the final examination. The data relating to their online activity with the related TELT resource was mapped alongside their examination score for the question, with results as follows:

- Average exam score overall for all students – 45%
- Not used error control codes TELT resource at all – 30%
- Used error control codes TELT during the module, at least once – 53%
- Used error control codes TELT specifically at revision May/June – 52%
- Used error control codes TELT on more than one specific date i.e. returned to package – 58 %

This snapshot relating to student performance suggests that those who engaged fully with the TELT resource did relatively well, although care should be taken to avoid confusion over correlation and causation (Ferguson and Clow, 2017).

From Figure 4 it can be seen that between weeks 9 and 20 about 200 students used the TELT resource relating to error control codes. Some of these might be students who used the resource more than once, so they are double counted in that figure of 200. Even so, although the use of the resource might be seen as disappointing, it would be reasonable to suppose that at least half the students used it in this period. That is also consistent with the 53% usage figure for students who attempted thexamination question relating to error control codes. However, far fewer than half of the student cohort felt confident to do the exam question. If students distributed themselves evenly across the three optional exam quesitons, we could expect 66% of them to attempt the question. It is reasonalble to suppose that by the time the students sat the examination they had not used the online resource for several months, unless they also used it in the revision period. From the A4A data it is possible to can see that very few students used it then. It is therefore reasonable to hypothesise that students shouldbe reminded of the importance of including the TELT resources in their revision. Engagement with the online materials should be encouraged, as they should enhance the learning experience by exposing students to a wider variety of learning techniques.

To help offset the limitations of small-scale research, the results for the focus sample in this study were compared with the wider cohort by reviewing student achievement relating to their cumulative Open University study on previous modules, as summarised in Figure 5.

Summary Statistics

Results

The MEANS Procedure

Analysis Variable : P_target_result1 Probability of student passing the module							
original_sample	N Obs	Mean	Std Dev	Minimum	Maximum	Mode	N
0	271	0.8670889	0.1349558	0.2279104	0.9876422	0.5742115	271
1	48	0.8516501	0.1430410	0.1702492	0.9850042	0.5742115	48

Figure 5 – statistical analysis of sample

To accomplish this, the student records for the focus sample were compared to all other student records from the cohort. The data set was coded as sample=1 for those who answered the particular exam question referred to in this study, and the others flagged sample = 0, so differentiating between the groups. The column p_target_result1 is the predicted probabilities of success for the students at module start. A comparison of the mean predictive probability in each group shows they are very similar although the sample group of 48 students is marginally weaker, as highlighted above. This suggests that even though the sample size is small there is no reason to assume that the sample group should perform any differently to the rest of the cohort.

4.2 Interview results

Two sets of interviews were conducted, one for the pilot study (3 students) and one for the following cohort (5 students). The Open University engages in distance learning so face-to-face interviews were not possible, as the students are widely distributed across the UK and beyond. For the pilot study the intial plan was for interviews to be conducted via the Open University’s Skype for Business system. However, during the interview period a restriction on recording external calls became evident, so interviews and recordings were completed via mobile phone. For the second cohort, interviews were conducted via Adobe Connect using its inbuilt recording facility. This had the added bonus of a visual screen, on which the online descriptions were added as an aide-memoir for the interviewees.

Several key points were highlighted by students during the interview process. For example, some students noted that the TELT resources were very good for self-testing. It was also noted that the activities provided a different way to learn rather than text and they were visual and interactive. As an example , the benefit of being able to step forwards and backwards through an animation so you can go back and check things was commented upon.

> “Seeing the coding in practice and having an interaction helped”.

A particular benefit relating to complex themes was noted, as the related TELT resources supplemented the wriiten text, thus helping the students to understand the topic more fully by interacting with the online version of the materials.

“ The sequence of coding and decoding is explained in the book and this is done quite well, but the activities help you to do that for real and allow you to apply the theory”.

It was also noted that the online resources were useful in providing a high-level summary, as large sections of the printed materials (2 or 3 pages) could be summarised in a paragraph or two of the online activity. This reinforces the suggestion of highlighting use of the TELT resources for revision, where students need to best utilise the time available. An issue noted was the lack of information regarding the estimated time for engaging with the online activities. The notional time needed varied between activities and this could not be determined unless actually engaging with the activity.

Although student perception of the TELT resources was generally positive there were also more negative responses, for example the following relating to the ‘Launching a wave’ activity.

“It was hard to understand the direction of the dipole and how it was radiating waves”.

The student found it difficult to work out what was happening. A slightly different animation showing which direction was which could have helped the student’s understanding, so this is a further idea to progress.

4.3 Discussion on recommendations and actions

Referring to Figure 2, the A4A framewok was adopted in order to focus on possible improvements relating to the student experience on TM355. The initial review of A4A data revealed that some students were using the online activites effectively to support their study of printed module materials, although many students did not fully engage. A particular issue was highlighted after the 2017 examination, so further analysis of the data was undertaken which suggested that those students who utilsed the TELT resource associated with the examination question performed slighlty better, although it should be noted that these results should be treated with some caution. Further action was taken by conducting a series of student interviews to gain further insight into their perception of the use of these activies.

The findings from this study suggest that there are several actions that could be taken, for example:

- Give a clearer indication of time needed for the TELT activities (although obviously this will vary for each student).
- Add short descriptions about what kind of activity it is, for example interactive, video.
- Promote the activities in a new module introductory or revision video or podcast.
- Use the module forums to promote them
- Have ‘talking heads’ of students saying how useful they were.
- Add further detail to the introduction to certain activities, for example to explain the orientation in the ‘launching a wave’ activity.

Several of these ideas suggested via the interviews have already been implemented and others could be actioned in the future. For example. Figure 6 depicts a section of a resource that has been produced to give students an overview of the activity type and typical timings, alongside a direct link to the activity and an indication on where it fits in the student study calendar.

Block	Part	Week	Online Activity	Activity Name and module link	Short Description	Estimated Time Req.
1	1	1-3	1.1	Fourier Transforms	Interactive activity showing time-domain representation and frequency-domain representation for sine, square, sawtooth and triangular waveforms.	15-30 mins

Figure 6 – TM355 TELT resource description and timing example

Also, a new revision podcast has been produced which specifically promotes the use of the TELT resource at revision time, hopefully resulting in more students revisiting the online resources.

5. Conclusions

Data analytics can prove useful in analysing student performance and in modifying a module in the light of what is revealed. However, as with any statistical data, interpretation is required. Data analytics do not 'speak for themselves'. Establishing the significance of analytics data is likely to require the use of additional strategies, of which interviews are an example. The example discussed here revealed some of the limitations of aggregated data. Knowing that a certain percentage of a student cohort did not do a particular activity could raise an alarm about the activity, but typically one would need to know more about the group of students identified. In what ways might they be representative or unrepresentative of the cohort? Pursuing this question is likely to require drilling down to data about individual students, which can be (as in the present case study) beyond what the analytics tool is intended to do, and might raise ethical concerns We see here another version of the 'prediction versus student experience' dilemma face by designers and users of data analytics tools.

In the present case study, follow-up interviews revealed the puzzling inconsistency that the TELT resources are considered useful but are under-utilised, particularly during the revision period. This fact indicates a clear course of action for the creators of the module, which is to urge students not to confine their revision solely to the printed texts, which contain the bulk of the teaching material. A revision advice podcast, newly introduced, stresses this and gives other revision advice.

Our general conclusion from this case study is that learning analytics have undoubtedly proved useful for tracking student engagement, but have required a certain amount of 'hand-crafting' to extract additional information that is not routinely available, and supplementary interviews to shed light on the potential significance of the data gleaned.

6. References

Bodily, R. & Verbert, K. (2017). Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. *IEEE Transactions on Learning Technologies*, 10(4), 405-418

Evans, G., Hidalgo, R. & Calder, K. (2017). Analytics for Action –enabling in-presentation evidence-based change on OU modules. *Quality Enhancement Report Series* Issue No 2017/7

Ferguson, R., & Clow, D. (2017, March). Where is the evidence? A call to action for learning analytics. In *LAK’17 Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 56–65). Vancouver, BC.

Gibson, D. & de Freitas, S. (2016). Exploratory analysis in learning analytics. *Technology, Knowledge and Learning*, 21 (1). pp. 5-19.

Kolb, D.A. (1984). *Experiential learning: experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice Hall.

Kop, R., Fournier, H. & Durand, G. (2017). A Critical Perspective on Learning Analytics and Educational Data Mining. In Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds) *Handbook of Leaning Analytics, Society for Learning Analytics Research*, 2017, pp.

Herodotou, C., Gilmour, A., Boroowa, A., Rienties, B., Zdrahal, Z. & Hlosta, M. (2017). Predictive modelling for addressing students’ attrition in Higher Education: The case of OU Analyse. In *CALRG Annual Conference 2017*, 14-16 Jun 2017, The Open University, Milton Keynes, UK.

Hidalgo, R. (2018). Analytics for action: Using data analytics to support students in improving their learning outcomes. In G. Ubachs & L. Konings Eds. *The envisioning report for empowering universities*(2nd ed), European Association of Distance Teaching Universities, Maastricht,Netherlands (pp. 6–8)

Littlejohn, A.(2017). Learning and Work: Professional Learning Analytics. In Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds) *Handbook of Leaning Analytics, Society for Learning Analytics Research*, 2017, pp 269-277

Macfadyen, L. P.& Dawson, S. (2010). Mining LMS Data to Develop an "Early Warning System" for Educators; a Proof of Concept. *Computers and Education*, 54, pp. 588-599

Sclater, N., Peasegood, A. & Mullen, J. (2016). *Learning Analytics in Higher Education, JISC report.* Retrieved from https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education

Shum, S. B., and Ferguson, R. (2012). Social learning analytics. *Educational Technology and Society*,15, 3–26

Siemens, G. (2019). Learning Analytics and Open, FLexible and Distance Learning. *Distance Education*, 40(3), 414-418

Slater S. & Baker R. (2019). Forecasting Student Master. *Distance Education*, 40(3), 380-394.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157e167. http://dx.doi.org/10.1016/

Tempelaar, D., Rienties, B. & Nguyen, Q. (2017). Towards actionable learning analytics using dispositions. *IEEE Transactions on Learning Technologies*, 10(1) pp. 6–16.

Viberg, O. Hatakka, M. Bälter & Mavroudi, A.(2018). The Current Landscape Analytics in Higher Education. *Computers in Human Behavior*, 2018, 98-110